

(19) 日本国特許庁 (J P)

## (12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-237991

(P2002-237991A)

(43) 公開日 平成14年8月23日 (2002.8.23)

(51) Int.Cl.<sup>7</sup>H 0 4 N 5/232  
7/15

識別記号

6 1 0  
6 3 0

F I

H 0 4 N 5/232  
7/15

テーマコード(参考)

Z 5 C 0 2 2  
6 1 0 5 C 0 6 4  
6 3 0 Z

審査請求 未請求 請求項の数3 O L (全 6 頁)

(21) 出願番号 特願2001-33351(P2001-33351)

(22) 出願日 平成13年2月9日(2001.2.9)

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72) 発明者 阿部 雅恵

東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内

(74) 代理人 100091546

弁理士 佐藤 正美

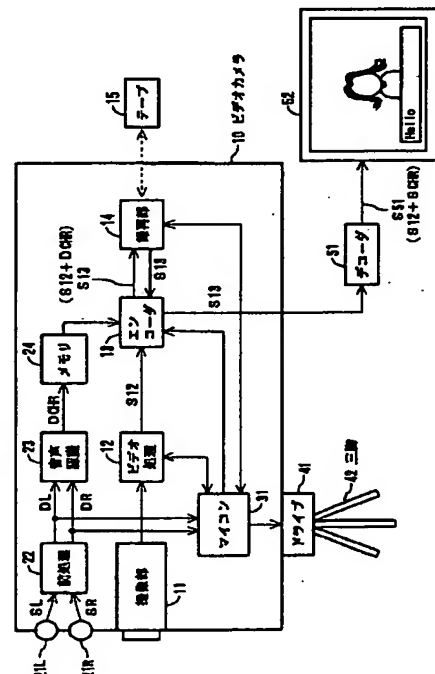
Fターム(参考) 5C022 AA12 AB62 AB63 AC01 AC27  
AC69 AC745C064 AA02 AB04 AC04 AC06 AC09  
AC13 AC16 AC22

(54) 【発明の名称】 ビデオカメラ

(57) 【要約】

【課題】 聴覚に障害のある者でも出席できるテレビ会議システムを提供する。

【解決手段】 撮像手段11と、この撮像手段11の撮像出力を標準のビデオ信号に変換する信号処理回路12とを設ける。発言者の発言を集音するためのマイクロフォン21L、21Rと、このマイクロフォン21L、21Rの出力信号に対して音声認識を行って対応する文字コードを出力する音声認識回路23とを設ける。ビデオ信号に、文字コードを重畳して出力するエンコーダ回路と、撮像手段11の向きを変更するドライブ機構41と、マイクロフォン21L、21Rの出力信号から発言者の位置を判別し、その判別結果をドライブ機構41に供給する制御回路31とを設ける。制御回路31は、判別結果により、撮像手段11が発言者を向くようにドライブ機構41を制御する。



1

## 【特許請求の範囲】

## 【請求項1】 撮像手段と、

この撮像手段の撮像出力を標準のビデオ信号に変換する  
信号処理回路と、

発言者の発言を集音するためのマイクロフォンと、

このマイクロフォンの出力信号に対して音声認識を行っ  
て対応する文字コードを出力する音声認識回路と、

上記ビデオ信号に、上記文字コードを重畳して出力する  
エンコーダ回路と、

上記撮像手段の向きを変更するドライブ機構と、

上記マイクロフォンの出力信号から発言者の位置を判別  
し、その判別結果を上記ドライブ機構に供給する制御回  
路とを有し、

上記制御回路は、上記判別結果により、上記撮像手段が  
上記発言者を向くように上記ドライブ機構を制御するよ  
うにしたビデオカメラ。

【請求項2】 請求項1に記載のビデオカメラにおいて、  
上記ビデオ信号に重畳された上記文字コードから、この  
文字コードに対応する文字を表示するビデオ信号を形成  
し、

この文字を表示するビデオ信号を上記ビデオ信号に加算  
して出力するデコーダを有するようにしたビデオカメ  
ラ。

【請求項3】 請求項1あるいは請求項2に記載のビデオ  
カメラにおいて、

上記マイクロフォンは1対とされ、

この1対のマイクロフォンの出力信号のレベル差から上  
記発言者の位置を判別するようにしたビデオカメラ。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 この発明は、テレビ会議シス  
テムなどに使用されるビデオカメラに関する。

## 【0002】

【従来の技術】 テレビ会議システムとして、会議室で行  
われている会議の様態をビデオカメラ（テレビカメラ）  
で撮影し、その内容を参考資料として磁気テープに記録  
したり、他の場所で観ることができるようにしたものがある。  
また、大規模なテレビ会議システムは、遠隔地にある複  
数の会議室を通信回線で結ぶことにより、会議の出席者  
は、互いの顔を見ながら双方向リアルタイムで会議を  
できるようにされている。

## 【0003】

【発明が解決しようとする課題】 しかし、上記のような  
テレビ会議システムは、聴覚に障害のある者が出席する  
会議には不適當である。

【0004】 この発明は、このような問題点を解決しよ  
うとするものである。

## 【0005】

【課題を解決するための手段】 この発明においては、例  
えば、撮像手段と、この撮像手段の撮像出力を標準のビ

2

デオ信号に変換する信号処理回路と、発言者の発言を集  
音するためのマイクロフォンと、このマイクロフォンの  
出力信号に対して音声認識を行って対応する文字コード  
を出力する音声認識回路と、上記ビデオ信号に、上記文  
字コードを重畳して出力するエンコーダ回路と、上記撮  
像手段の向きを変更するドライブ機構と、上記マイクロ  
フォンの出力信号から発言者の位置を判別し、その判別  
結果を上記ドライブ機構に供給する制御回路とを有し、  
上記制御回路は、上記判別結果により、上記撮像手段が  
上記発言者を向くように上記ドライブ機構を制御するよ  
うにしたビデオカメラとするものである。したがって、  
複数の会議出席者のうち、発言者がビデオカメラに自動  
的に撮像されるとともに、その発言内容が字幕により表  
示される。

## 【0006】

【発明の実施の形態】 図1は、この発明によるビデオカ  
メラをテレビ会議で使用している状態の一例を示し、図  
1においては、4人の出席者A、B、C、Dにより会議  
が行われる場合である。

【0007】 そして、これら出席者A～Dの前に、この  
発明によるビデオカメラ10が配置されている。このビ  
デオカメラ10は、ドライブ機構41を通じて三脚42  
に取り付けられ、ドライブ機構41により水平方向にパ  
ンできるようにされている。また、詳細は後述するが、  
ビデオカメラ10から出力されるビデオ信号が、デコー  
ダ51を通じてモニタ受信機52に供給される。したが  
って、ビデオカメラ10は出席者A～Dの誰でも撮像で  
き、その画像をモニタ受信機52に表示することができる。

【0008】 そして、ビデオカメラ10は、例えば図2  
に示すように構成されている。すなわち、符号11はそ  
の撮像部を示し、この撮像部11は、撮像レンズおよび  
CCD撮像素子などを有し、この撮像部11により出席  
者A～Dが選択的に撮像される。

【0009】 また、例えば図1にも示すように、ビデオ  
カメラ10の前部上方には、1対のマイクロフォン21  
L、21Rが設けられる。このマイクロフォン21L、  
21Rは、例えば単一指向性のものとされるときに、  
それらの指向軸が、同一の水平面内において、ビデオカ  
メラ10の撮像中心（光軸）に対して対称で、かつ、所  
定の角間隔を有するように、取り付けられている。

【0010】 そして、会議のときには、マイクロフォン  
21L、21Rの出力信号SL、SRが、前処理回路22  
に供給されて各種の処理、例えば以後の処理に必要な帯  
域成分だけが取り出されるときに、デジタルオーディ  
オデータDL、DRにA/D変換され、これらデジタルオー  
ディオデータDL、DRが制御用のマイクロコンピュー  
タ31に供給されてデジタルオーディオデータDL、DR  
となった音声の発言者（音源）の方向が判別される。な  
お、この判別方法の一例を後述する。

3

【0011】そして、この判別結果がドライブ機構41に供給され、撮像部11の撮像中心が、デジタルオーディオデータDL、DRとなった音声の発言者に向くように、ビデオカメラ10の向きが制御される。したがって、ビデオカメラ10は発言者を自動的に捕捉して追尾することになる。また、この捕捉および追尾により、音声信号SL、SRは、主としてその発言者のものとなる。

【0012】さらに、撮像部11の撮像出力がビデオ信号処理回路12に供給されて標準のビデオ信号、例えばNTSC方式のビデオ信号S12に変換され、このビデオ信号S12がエンコード回路13に供給される。

【0013】また、前処理回路22からのデジタルオーディオデータDL、DRが音声認識回路23に供給されて音声認識され、マイクロフォン21L、21Rの集音した音声、対応する文字コード(テキストデータ)DCHRに変換され、この文字コードDCHRがメモリ回路24においてタイミングが調整されてからエンコード回路13に供給される。

【0014】このエンコード回路13は、クローズド・キャプションと呼ばれるテレビ文字多重放送(テレテキスト)におけるエンコード処理を実行するものであり、ビデオ処理回路12からのビデオ信号S12の第21番目および第284番目の水平走査期間に、文字コードDCHRを重ねさせるものである。したがって、エンコード回路13からは、ビデオ信号S12に文字コードDCHRの重畳されたビデオS13が出力される。

【0015】そして、このビデオ信号S13が記録再生部14に供給される。この記録再生部14は、一般のVTRと同様に構成されているものであり、供給されたビデオ信号S13は、所定の記録信号(FM輝度信号と低域変換搬送色信号との周波数多重化信号)に変換され、回転磁気ヘッドによりカセットテープ15の磁気テープに記録される。

【0016】また、エンコード回路13からのビデオ信号S13が、デコード51に供給される。このデコード51は、エンコード回路13とは相補の処理を行うものであり、これに供給されたビデオ信号S13は、もとのビデオ信号S12と、文字コードDCHRとに分離されるとともに、その文字コードDCHRが対応する文字を表示するビデオ信号SCHRに変換され、この文字のビデオ信号SCHRがビデオ信号S12に加算される。こうして、デコード51からは、ビデオ信号S12に、文字のビデオ信号SCHRの加算されたビデオ信号S51が取り出される。そして、このビデオ信号S51が、モニタ受像機52に供給される。

【0017】したがって、会議のとき、出席者A~Dのうちの誰か、例えば出席者Cが発言すると、その音声にしたがって捕捉および追尾が行われ、ビデオカメラ10は発言者Cに向くようになり、発言者Cを撮像する。この結果、図1にも示すように、モニタ受信機52には、

4

発言者Cが表示されることになる。

【0018】また、このとき、音声認識回路23において、発言者Cの発言内容が音声認識されるとともに、デコード51において、その認識結果を文字として表示するビデオ信号SCHRが形成されてモニタ受像機52に供給されるので、図1にも示すように、モニタ受像機52には、発言者Cと同時に、その発言内容が字幕として表示される。

【0019】一方、会議の模様を記録したカセットテープ15を記録再生部14にセットして再生の操作を行うと、記録再生部14において、記録時とは相補の処理が実行されてビデオ信号S13が再生され、このビデオ信号S13が、記録再生部14からエンコード回路13を通じてデコード51に供給されてビデオ信号S51が形成される。そして、このビデオ信号S51がモニタ受像機52に供給される。なお、この再生時には、エンコード回路13におけるエンコード処理は実行されない。

【0020】したがって、モニタ受像機52には、会議のときに撮像された内容、すなわち、発言者が表示されるとともに、その発言内容が字幕として表示される。

【0021】こうして、上述のシステムによれば、会議における発言者が自動的に捕捉および追尾されてモニタ受像機52に表示されるとともに、その発言内容が字幕として表示されるので、聴覚に障害のある者でも、その会議に出席することができる。また、テープ15に記録された会議の内容も、発言者とともに、字幕付きで表示することができるので、その会議の内容を事後に知ることでもできる。

【0022】さらに、発言者の撮像や字幕の表示は自動的に行われ、そのための人手を必要としない。

【0023】図3は、ビデオカメラ10に発言者を捕捉および追尾させるためのフローチャートを示し、符号100はそのルーチンである。このルーチン100は、マイクロコンピュータ31に用意され、ビデオ信号S12の例えば1フィールド期間ごとにマイクロコンピュータ31により実行される。

【0024】すなわち、ルーチン100においては、マイクロコンピュータ31の処理がステップ101からスタートし、次にステップ102において、前処理回路22から出力されるデジタルオーディオデータDL、DR(時点 $t_i$ におけるデータDL、DR)がマイクロコンピュータ31に取り込まれる。

【0025】そして、ステップ103において、データDLとデータDRとの差のデータ $D_i$ が計算される。この場合、データDL、DRは、時点 $t_i$ にマイクロフォン21L、21Rの集音した音声の信号であるから、データ $D_i$ は、時点 $t_i$ にマイクロフォン21L、21Rの集音した音声の音量差を示すことになる。また、ルーチン100はフィールド周期で実行されるので、データ $D_i$ はフィールド期間ごとに得られることになるが、その最新

5

の60フィールド分のデータ $D_i$  ( $i=1\sim60$ )、つまり、最新の1秒分のデータ $D_1\sim D_{60}$ が常にマイクロコンピュータ31にいったん保持される。

【0026】次にステップ104において、最新の60フィールド分のデータ $D_1\sim D_{60}$ が加算されて和(積分値)のデータ $DSUM$ が求められる。この場合、データ $D_1\sim D_{60}$ は、それぞれ時点 $t_1\sim t_{60}$ にマイクロフォン21L、21Rの集音した音声の音量差であるから、和のデータ $DSUM$ は、最新の1秒間におけるマイクロフォン21L、21Rの集音した音声の音量差を示すことになる。

【0027】続いてステップ111において、データ $DSUM$ の絶対値 $|DSUM|$ が所定値 $DTH$ と比較され、 $|DSUM| \geq DTH$ のときには、処理はステップ111からステップ112に進み、このステップ112において、データ $DSUM$ の符号が判別される。

【0028】そして、この判別の結果、 $DSUM > 0$ ならば、最新の1秒間においては、マイクロフォン21Lの集音した音声の音量が、マイクロフォン21Rの集音した音声の音量よりも大きいのであるから、ビデオカメラ10から見て左側に音源があると考えられる。すなわち、発言者が左側にいるとみなせる。逆に、 $DSUM < 0$ ならば、最新の1秒間においては、マイクロフォン21Rの集音した音声の音量が、マイクロフォン21Lの集音した音声の音量よりも大いのであるから、ビデオカメラ10から見て右側に発言者がいるとみなせる。

【0029】そこで、 $DSUM > 0$ のときには、すなわち、発言者がカメラ10の左側にいるときには、処理はステップ112からステップ113に進み、このステップ113において、ドライブ機構41に所定のドライブ電圧が供給され、ビデオカメラ10は所定の角度だけ左方向に向けられ、その後、ステップ115によりルーチン100を終了する。

【0030】そして、発言者がビデオカメラ10の左側にいる間は、以上のステップ101~104、111~113、115が繰り返されるので、ビデオカメラ10は次第に左方向にパンしていき、発言者を画面内に捕らえるようになる。

【0031】また、 $DSUM < 0$ のときには、発言者がカメラ10の右側にいるときであるが、このときには、処理はステップ112からステップ114に進み、このステップ114において、ドライブ機構41に所定のドライブ電圧が供給され、ビデオカメラ10は所定の角度だけ右方向に向けられ、その後、ステップ115によりルーチン100を終了する。

6

【0032】そして、発言者がビデオカメラ10の右側にいる間は、以上のステップ101~104、111、112、114、115が繰り返されるので、ビデオカメラ10は次第に右方向にパンしていき、発言者を画面内に捕らえるようになる。

【0033】そして、ビデオカメラ10が左方向あるいは右方向にパンしていくことにより、発言者が画面の中央に位置するようになると、このとき、マイクロフォン21L、21Rの集音する音声の音量はほぼ等しくなり、その結果、 $DL \approx DR$ となって $|DSUM| < DTH$ となる。すると、これが、ステップ111において判別され、処理はステップ111からステップ115に進み、ビデオカメラ10をパンさせることなくルーチン100を終了する。したがって、以後、発言者はビデオカメラ10により連続して撮像されることになる。

【0034】こうして、ルーチン100によれば、発言者に対して自動的に補捉および追尾が行われ、常に発言者を撮像することができる。

【0035】〔この明細書で使用している略語の一覧〕

CCD : Charge Coupled Device

FM : Frequency Modulation

NTSC : National Television System Committee

VTR : Video Tape Recorder

【0036】

【発明の効果】この発明によれば、会議における発言者がモニタ受像機に表示されるとともに、その発言内容が字幕として表示されるので、聴覚に障害のある者でも、その会議に出席することができる。また、カセットテープに記録された会議の内容も、発言者とともに、字幕付きで表示することができるので、その会議の内容を事後に知ることもできる。

【図面の簡単な説明】

【図1】この発明の使用形態を示す図である。

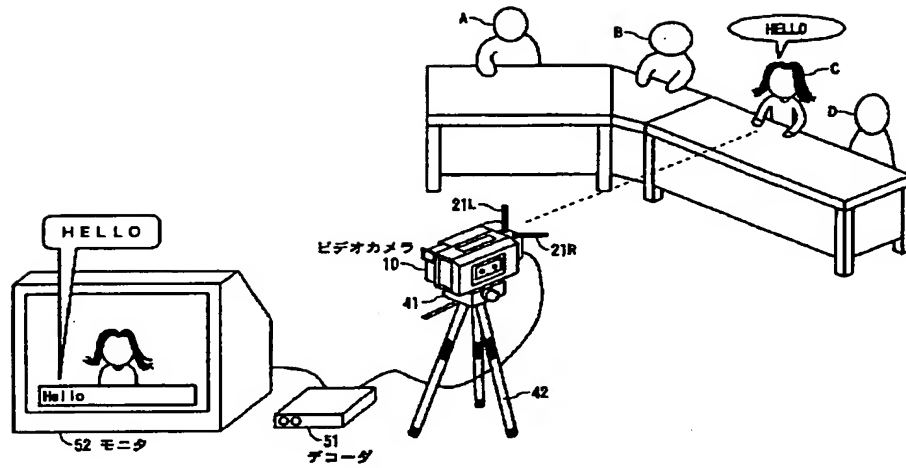
【図2】この発明の一形態を示す系統図である。

【図3】この発明の一形態を示すフローチャートである。

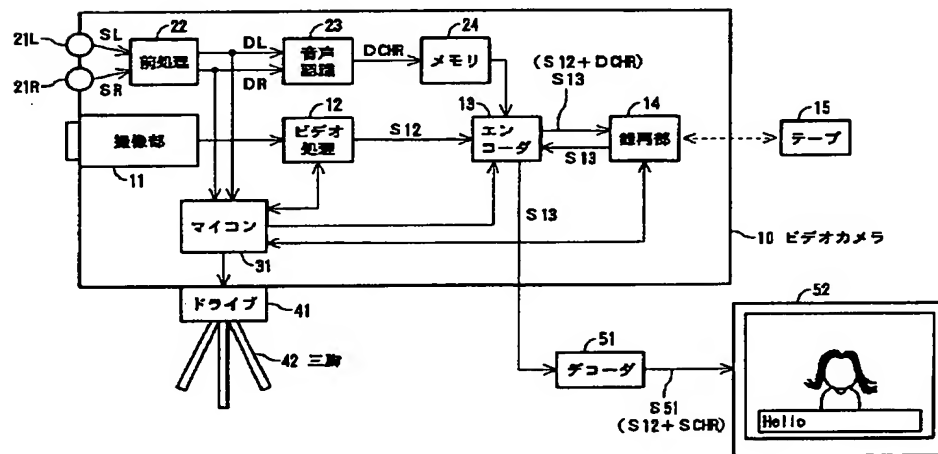
【符号の説明】

10…ビデオカメラ、11…撮像部、12…ビデオ信号処理回路、13…エンコーダ回路、14…記録再生部、15…カセットテープ、21Lおよび21R…マイクロフォン、22…前処理回路、23…音声認識回路、24…メモリ回路、31…マイクロフォン、41…ドライブ機構、42…三脚、51…デコーダ、52…モニタ受像機

【図 1】



【図 2】



【図3】

